



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

AssayR – a simple mass spectrometry software tool for targeted metabolic and stable isotope tracer analyses.

Citation for published version:

Wills, J, Edwards-hicks, J & Finch, A 2017, 'AssayR – a simple mass spectrometry software tool for targeted metabolic and stable isotope tracer analyses.', *Analytical Chemistry*.
<https://doi.org/10.1021/acs.analchem.7b02401>

Digital Object Identifier (DOI):

[10.1021/acs.analchem.7b02401](https://doi.org/10.1021/acs.analchem.7b02401)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Analytical Chemistry

Publisher Rights Statement:

ACS AuthorChoice - This is an open access article published under a Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

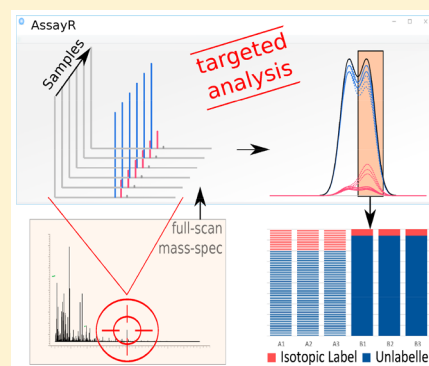


AssayR: A Simple Mass Spectrometry Software Tool for Targeted Metabolic and Stable Isotope Tracer Analyses

Jimi Wills,* Joy Edwards-Hicks, and Andrew J. Finch*

Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XR, United Kingdom

ABSTRACT: Metabolic analyses generally fall into two classes: unbiased metabolomic analyses and analyses that are targeted toward specific metabolites. Both techniques have been revolutionized by the advent of mass spectrometers with detectors that afford high mass accuracy and resolution, such as time-of-flights (TOFs) and Orbitraps. One particular area where this technology is key is in the field of metabolic flux analysis because the resolution of these spectrometers allows for discrimination between ^{13}C -containing isotopologues and those containing ^{15}N or other isotopes. While XCMS-based software is freely available for untargeted analysis of mass spectrometric data sets, it does not always identify metabolites of interest in a targeted assay. Furthermore, there is a paucity of vendor-independent software that deals with targeted analyses of metabolites and of isotopologues in particular. Here, we present AssayR, an R package that takes high resolution wide-scan liquid chromatography–mass spectrometry (LC-MS) data sets and tailors peak detection for each metabolite through a simple, iterative user interface. It automatically integrates peak areas for all isotopologues and outputs extracted ion chromatograms (EICs), absolute and relative stacked bar charts for all isotopologues, and a .csv data file. We demonstrate several examples where AssayR provides more accurate and robust quantitation than XCMS, and we propose that tailored peak detection should be the preferred approach for targeted assays. In summary, AssayR provides easy and robust targeted metabolite and stable isotope analyses on wide-scan data sets from high resolution mass spectrometers.



The goal of an untargeted metabolomic experiment is usually to identify metabolites that have changed with the greatest significance or magnitude between two or more experimental conditions. A typical untargeted mass spectrometric experiment usually follows a well-defined workflow, using proprietary or open source software (e.g., XCMS,^{1–3} mzMine^{4,5}) to give a list of features that can be quantified and matched to a database to yield probable or verified metabolite identifications. This approach was recently extended to include untargeted identification of stable isotope fluxes using the elegant X13CMS software tool.⁶ In contrast, a targeted metabolite experiment is one in which specific metabolites must be identified with high confidence in all samples (where detectable), and this requires prioritization of different analytical parameters. Existing targeted workflows based upon XCMS do exist,⁷ but the enforcement of a single set of global peak detection parameters is a limitation that can lead to missed peaks or inaccurate quantitation. Some peaks are simply not found, particularly with mixed mode hydrophilic liquid interaction (HILIC) chromatography where peaks can be broad and of irregular shape. Furthermore, this approach suffers serious limitation in the analysis of stable isotope tracing experiments because isotopologues are treated as distinct features during the peak detection stage when they should be detected in concert. This also impacts upon data output, since isotopologues are not grouped together and must therefore be

further processed to yield the isotopic composition of each metabolite.

Targeted metabolic analysis has traditionally required less postacquisition analysis because the preferred instrument for such experiments has been the triple quadrupole mass spectrometer, and the combination of precursor and product m/z ions specified at the point of data acquisition is tied to a specific metabolite.⁸ With this strategy, metabolite identification is primarily a preacquisition issue rather than a postacquisition one. Adding a metabolite tracer into such an analysis, however, necessitates the addition of MRM (multiple reaction monitoring) transitions for each expected isotopologue, and this yields a complexity of acquisition that is not desirable, quickly limiting the number of metabolites that can be measured. The problem of acquisition complexity is even more pronounced if isotopic tracers are used that contain more than one heavy isotope (e.g., $^{13}\text{C}_5$, $^{15}\text{N}_2$ -glutamine). It is in this context that the new generation of high resolution, accurate mass spectrometers excel because relatively standard wide scan methods can be used for data acquisition, yet many metabolites and their isotopologues can subsequently be separated and quantified through data analysis approaches.⁹

Received: June 21, 2017

Accepted: August 29, 2017

Published: August 29, 2017

We set several criteria for an ideal software tool that can take high resolution, high mass accuracy data from any mass spectrometer and return peak integrals for specific metabolites and their isotopologues. These criteria are (a) robust peak detection taking into account all isotopologues, (b) a simple, optional quality control curation step for all peaks prior to quantitation, (c) reporting of values for separate (including split) peaks where more than one is found that could be attributed to a single metabolite, (d) reporting of values and bar charts for grouped isotopologues, and (e) an interface that is easy and intuitive to use. Here, we present AssayR, an R package¹⁰ that fulfills the above criteria (Figure 1). Using data

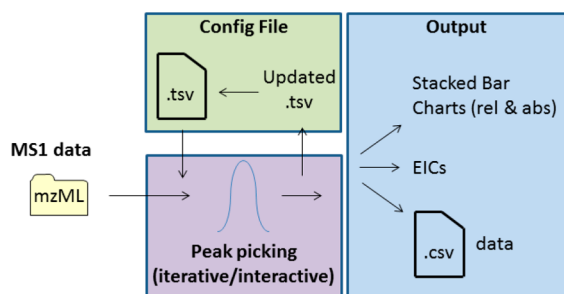


Figure 1. Schematic of AssayR showing the main concepts and demonstrating minimal user input (initial config and optional peak picking only). mzML files undergo extracted ion chromatogram (EIC) analysis based upon the m/z values in the input config file. Optional interactive peak picking leads to a final config file which is used to produce the peak integrals for quantitation. All required isotopologues are included in the process, and the outputs are a .csv file of the data as well as EICs and bar charts of relative (percentage) and absolute values for all isotopologues.

obtained on a ThermoScientific Q Exactive mass spectrometer, we demonstrate outputs from XCMS and AssayR that reveal more accurate and robust quantitation of analytes in AssayR.

METHODS

Analysis of Cellular Metabolites. MRC5 primary human fibroblasts were switched to DMEM with 25 mM $^{13}\text{C}_6$ -glucose for 5 or 60 min. The medium was aspirated; cells were washed quickly with ice-cold PBS, and metabolites were extracted with 50:30:20 methanol/acetonitrile/water. Samples (triplicates) were applied to liquid chromatography–mass spectrometry (LC-MS) using a 15 cm \times 4.6 mm ZIC-pHILIC (Merck Millipore) column fitted with a guard on a Thermo Ultimate 3000 HPLC. A gradient of decreasing acetonitrile (with 20 mM ammonium carbonate as the aqueous phase) was used to elute metabolites into a Q Exactive mass spectrometer. Data were acquired in wide scan negative mode. In order to generate mzML files, the command “msconvert all()” was run that uses the msconvert utility of Proteowizard^{11,12} to generate separate positive and negative mode mzML files.

SOFTWARE DESCRIPTION

Input File Format. AssayR uses the R package mzR to extract chromatograms from files in mzML format.

Config File. A configuration file in .tsv format is associated with each analysis (Figure 2). This file specifies the m/z value and the retention time (RT) window of each metabolite of interest as well as the maximum number of isotopologues to analyze (split into ^{13}C , ^{15}N , and ^2H). For config file setup purposes, the full retention time range can be selected (e.g.,

Initial config file

name	mz	ppm	rt.min	rt.max	seconds	threshold	C13	N15	H2	interactive
Glucose	179.0561	5	1	40	20	1.00E+05	6	0	0	yes
Pyruvate	87.0087	5	1	40	20	1.00E+05	3	0	0	yes

Interactive peak picking

Final config file

name	mz	ppm	rt.min	rt.max	seconds	threshold	C13	N15	H2	interactive
Glucose	179.0561	5	14	16.5	20	1.00E+06	6	0	0	yes
Pyruvate	87.0087	5	9.5	10.8	11	1.00E+06	3	0	0	yes

Figure 2. Examples of Initial and Final config files. Typical default values are given in the Initial config file. “seconds” refers to the width of the peak detection filter and not the peak width. The red box highlights parameters that are modified during interactive peak picking. The blue box highlights the simple isotopologue number input.

Initial config file in Figure 2) as well as default values for the width of the peak detection filter (“seconds”; see [Extracted Ion Chromatogram Generation and Peak Detection](#) below) and intensity threshold. An “interactive” option is also included so that the user can opt out of the iterative peak detection step for any metabolite, for instance, if it is known that the peak is always picked correctly by the current settings. Isotopologue selection is simply a numerical input for ^{13}C , ^{15}N , or ^2H , and combined isotopes can be selected: all possible isotopologues are analyzed.

Extracted Ion Chromatogram Generation and Peak Detection. A more detailed description accompanies the R package code, which is available at <https://gitlab.com/jimiwillis/assay.R>. Briefly, a row from the configuration table (Figure 2), representing an analyte, is read and the configured m/z ranges (combining m/z , ppm, and isotope settings) are extracted from mzML files via the mzR package. Interpolation is used to standardize the retention times across these chromatograms, and the maximal chromatographic profile is taken forward for peak detection. This means that a peak only needs to be present in a single sample for a single isotope for that peak to be detected and measured across the whole context. The use of combined isotopologues (Figure 3A) for metabolite peak identification is particularly important when a mix of labeled and unlabeled samples are analyzed or for samples where the labeling in a given metabolite is saturated, and therefore, the monoisotopic m/z value would be inappropriate for metabolite peak identification.

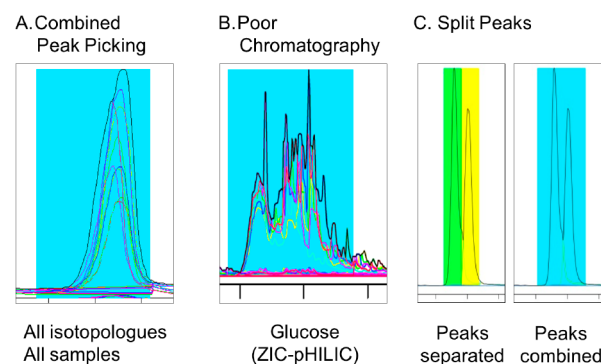


Figure 3. Peak detection in AssayR. (A) Peak detection (shaded blue) is specified for each metabolite based upon all isotopologues in all samples. (B) Example of peak detection (blue shading) despite poor chromatography. (C) AssayR enables split peaks to be detected separately (shaded green/yellow) or together. Shaded areas are detected and quantified.

The retention time minimum and maximum from the configuration table are used to select a region of the chromatogram on which peak detection will be performed. A Mexican hat filter is used with the filter function in R to translate the chromatogram and to set start and end indices for each detected peak. The indices are used to define the range to be summed to generate peak area measurements for each chromatogram. Individual peaks are marked (in blue), where more than one peak is identified within a metabolite retention time window; the peaks are separated and shaded in different colors (e.g., green/yellow for 2 peaks; Figure 3C), to assist with peak curation. The interactive peak picking procedure then allows simple alteration of detection parameters through an intuitive query-based format. Alteration of the width of the Mexican hat ("seconds" column in the config file) enables most peaks to be picked, even when the chromatography is poor, such as when measuring glucose on a ZIC-pHILIC column (Figure 3B). Split peaks can also be selected as a single metabolite or two peaks by alteration of the hat width (Figure 3C). The minimum and maximum m/z values, the hat filter width, and the peak detection threshold are all updated during the interactive peak picking process. Once the user is satisfied with the result, the updated parameters are written back to the configuration file (e.g., Final config file in Figure 2) and the peak areas are saved for output in the data .csv file.

Output. The primary output from AssayR is a .csv file with samples separated by column and metabolites/isotopologues separated by row. Images of the extracted ion chromatograms generated during peak detection are exported for recording metabolite identification and quality control: these images are generated even when the software is run without interactive peak curation. Stacked bar charts of absolute and relative peak intensity are produced for each metabolite, allowing quick and easy visualization of the data. These reveal variance between samples and allow for quick identification of possible outlier samples (as outlined below). A representative analysis of $^{13}\text{C}_6$ -glucose tracing in primary human fibroblasts is presented in Figure 4.

■ COMPARISON WITH XCMS

A popular chromatographic approach in metabolomics is the use of ZIC-pHILIC columns at high pH^9 because they capture a wide range of metabolites, including most of the organic acids of central carbon metabolism. However, the chromatographic performance of these matrices can be poor, especially in comparison to reversed-phase chromatographic approaches. Variability in peak shape can be pronounced as some metabolites can interact with the matrix in more than one way, and this can lead to spread (e.g., glucose in Figure 3B) or separated peaks. This variability can be more pronounced if methanol is used during sample loading due to additional surface effects of the solvent. Using the metabolomic data set from MRC5 primary human fibroblasts pulsed with $^{13}\text{C}_6$ -glucose, we analyzed glycolytic and related metabolites with AssayR (Figure 4) and XCMS. As described above, the chromatography of glucose is poor and XCMS did not pick any of the isotopologues, whereas AssayR showed almost full labeling in all samples (Figure 5A). While fructose 6-phosphate was well resolved and accurately picked by both packages, the monoisotopic glucose 6-phosphate peak had an overlapping isobaric peak with slightly later retention time (Figure 5B). During the peak detection stage in AssayR, it was clear that these were mixed metabolites because some of the extracted ion

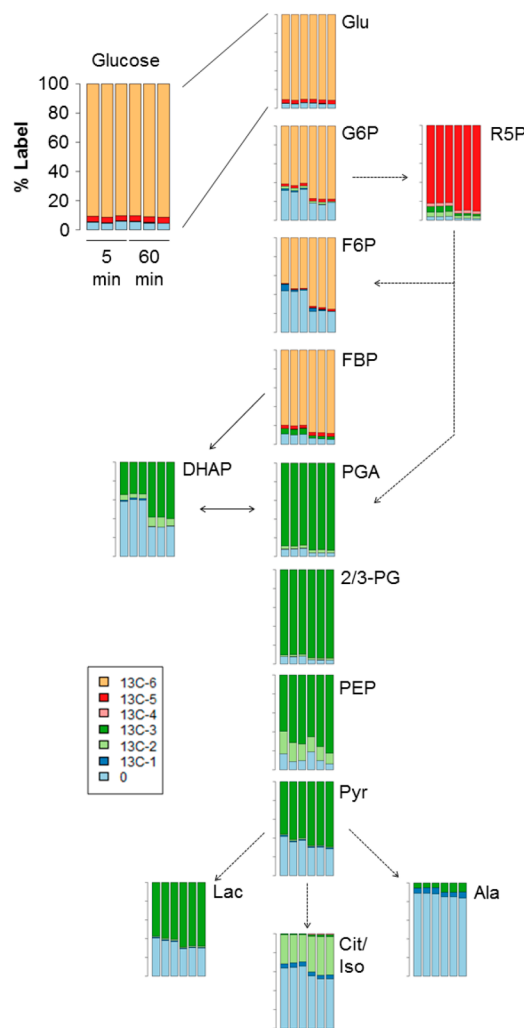


Figure 4. Glycolytic and related stable isotope tracing of $^{13}\text{C}_6$ -glucose metabolism quantified by AssayR. Relative (percentage) stacked bar charts of triplicates are shown as produced by AssayR (absolute stacked bar charts and EICs are also produced automatically). MRC-5 fibroblasts were pulsed for 5 or 60 min with $^{13}\text{C}_6$ -glucose in triplicate. Abbreviations: Glu (glucose), G6P (glucose 6-phosphate), F6P (fructose 6-phosphate), FBP (fructose 1,6-bisphosphate), PGA (3-phosphoglycerate), DHAP (dihydroxyacetone phosphate), 2/3-PG (2-/3-phosphoglycerate), PEP (phosphoenolpyruvate), Pyr (pyruvate), Ala (alanine), Lac (lactate), Cit/Iso (citrate/isocitrate).

chromatograms (EICs) matched the first peak only whereas some matched both peaks (Figure 5B). AssayR was set up to resolve these peaks, but XCMS picked them together.

The stacked bar plots in AssayR revealed that the later peak was unlabeled, whereas the earlier peak (glucose 6-phosphate) was predominantly labeled. Due to the analysis of mixed metabolites, XCMS underestimated the labeling of glucose 6-phosphate and gave a high variance (Figure 5B). A third problem was noticed with the quantitation of ^{13}C incorporation into citrate/isocitrate. Comparison of the stacked bar plots revealed that the third sample in the 60 min time point (sample 6) was different from the other two, showing lower $^{13}\text{C}_2$ abundance (Figure 5C). Examination of the individual EICs and RT over which they were integrated revealed that XCMS had only picked part of the peak in sample 6 (red area of EIC), and therefore, this isotopologue was underrepresented in the analysis. This type of error cannot occur in AssayR because

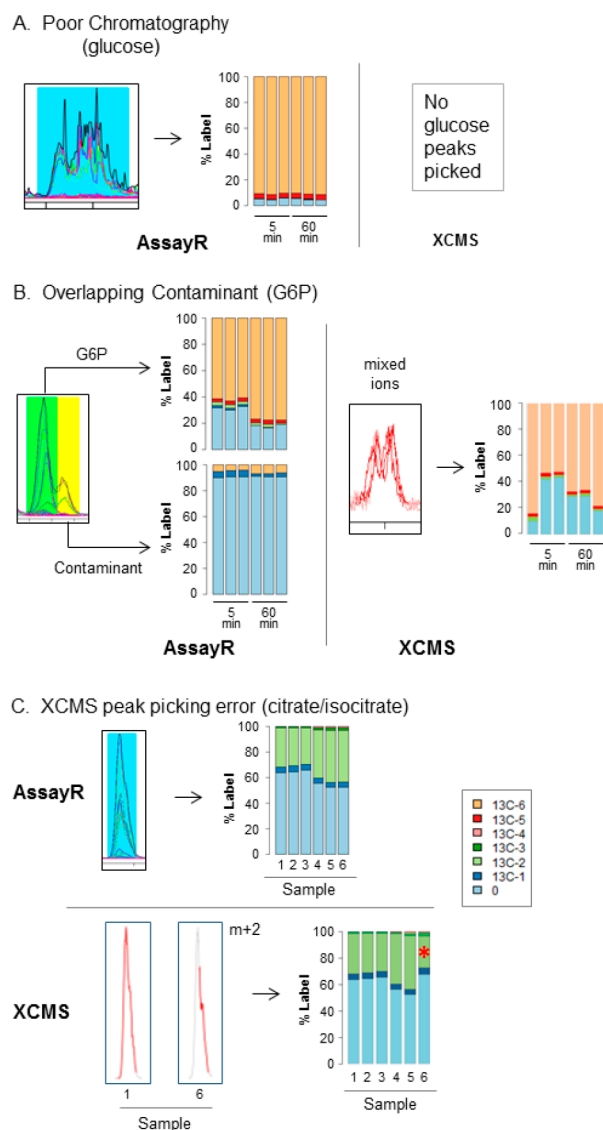


Figure 5. Comparison of AssayR with XCMS. (A) Peaks that fail the XCMS peak detection are picked and quantified with AssayR. (B) User control over peak detection in AssayR allows exclusion of incorrect peaks, particularly with overlapping isobaric species (the chromatograms are different because AssayR includes all isotopologues; the monoisotopic only is shown for XCMS). (C) Example of misquantitation by XCMS due to partial peak detection of a ($m + 2$) isotopologue. XCMS detected/quantified area of the EIC is in red (red asterisk indicates inaccurate $m + 2$ quantitation in the corresponding bar chart). G6P = glucose 6-phosphate.

integration occurs over a fixed RT window for all isotopologues. Thus, we present data that strongly support the use of tailored peak detection for the quantitation of specific metabolites in wide scan high resolution LC-MS data sets.

CONCLUSION

AssayR is an open source platform-agnostic R package that enables straightforward analysis of high resolution mass spectrometric data sets for targeted analyses, particularly those involving stable isotope tracers. The increasing availability of high resolution mass spectrometers renders this a timely addition to the analytical capability of investigators studying metabolic pathways. While common preference for the

reliability and quantitative capability of triple-quadrupole mass spectrometers will not be displaced in the immediate future by high resolution spectrometers, the versatility of the post-acquisition approach afforded by the latter is a very good match for stable isotope labeling studies. AssayR enables a simple, robust, and powerful approach to the measurement of metabolite usage in biological samples.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jimi.wills@igmm.ed.ac.uk

*E-mail: andrew.finch@igmm.ed.ac.uk

ORCID

Andrew J. Finch: 0000-0002-8065-4623

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

AssayR, an example config file, and the data set for this article may be accessed at no charge on the Gitlab repository at <https://gitlab.com/jimiwills/assayR/>.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council and the Wellcome Trust Institutional Support Fund.

REFERENCES

- (1) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, 78 (3), 779–787.
- (2) Tautenhahn, R.; Bottcher, C.; Neumann, S. *BMC Bioinf.* **2008**, 9 (1), 504.
- (3) Benton, H. P.; Want, E. J.; Ebbels, T. M. D. *Bioinformatics* **2010**, 26 (19), 2488–2489.
- (4) Katajamaa, M.; Miettinen, J.; Oresic, M. *Bioinformatics* **2006**, 22 (5), 634–636.
- (5) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. *BMC Bioinf.* **2010**, 11 (1), 395.
- (6) Huang, X.; Chen, Y.-J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J. *Anal. Chem.* **2014**, 86 (3), 1632–1639.
- (7) Creek, D. J.; Jankevics, A.; Burgess, K. E. V.; Breitling, R.; Barrett, M. P. *Bioinformatics* **2012**, 28 (7), 1048–1049.
- (8) Lu, W.; Bennett, B. D.; Rabinowitz, J. D. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, 871 (2), 236–242.
- (9) Mackay, G. M.; Zheng, L.; van den Broek, N. J. F.; Gottlieb, E. *Methods Enzymol.* **2015**, 561, 171–196.
- (10) R Foundation. *The R Project for Statistical Computing*; The R Foundation: Vienna, Austria; ISBN: 3-900051-07-0; www.R-project.org [Accessed 08/24/2017].
- (11) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, 24 (21), 2534–2536.
- (12) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat. Biotechnol.* **2012**, 30 (10), 918–920.